

## Aproximación metodológica al uso de redes neuronales artificiales para la predicción de resultados en medicina

Javier Trujillano, Jaume March y Albert Sorribas

Grup de Recerca de Biomatemàtica i Bioestadística. Departament de Ciències Mèdiques Bàsiques. Universitat de Lleida. España.

En la práctica clínica, existe un gran interés en conseguir modelos de predicción precisos. Dentro de las posibles alternativas técnicas disponibles, las redes neuronales artificiales (RN) se utilizan cada vez más en medicina. En esta revisión realizamos primero una introducción a la metodología de RN, describiendo el funcionamiento del tipo de RN más utilizado, que es el perceptrón multicapa entrenado con algoritmo de retropropagación (MLP). A continuación, comparamos el MLP con la regresión logística (RL) y, finalmente, mediante un ejemplo con datos reales, mostramos un esquema práctico para realizar una aplicación basada en RN. La ventaja principal de las RN es su capacidad de incorporar efectos no-lineales e interacciones entre las variables del modelo sin necesidad de incluirlas a priori. Como mayores desventajas, las RN proporcionan un modelo donde es difícil la interpretación de los parámetros y que necesita de un cierto empirismo en su proceso de construcción y entrenamiento. Las RN son útiles para el cálculo de probabilidad de un resultado en función de un conjunto de variables predictoras y consiguen mejores resultados, en algunas situaciones, que la RL. Las metodologías de RN y RL son complementarias y nos ayudan a conseguir modelos más válidos.

*Palabras clave:* Red neuronal artificial. Predicción de resultados. Regresión logística.

Methodological approach to the use of artificial neural networks for predicting results in medicine

In clinical practice, there is an increasing interest in obtaining adequate models of prediction. Within the possible available alternatives, the artificial neural networks (ANN) are progressively more used. In this review we first introduce the ANN methodology, describing the most common type of ANN, the Multilayer Perceptron trained with backpropagation algorithm (MLP). Then we compare the MLP with the Logistic Regression (LR). Finally, we show a practical scheme to make an application based on ANN by means of an example with actual data. The main advantage of the RN is its capacity to incorporate nonlinear effects and interactions between the variables of the model without need to include them a priori. As greater disadvantages, they show a difficult interpretation of their parameters and large empiricism in their process of construction and training. ANN are useful for the computation of probabilities of a given outcome based on a set of predicting variables. Furthermore, in some cases, they obtain better results than LR. Both methodologies, ANN and LR, are complementary and they help us to obtain more valid models.

*Key words:* Artificial neural networks. Outcome prediction. Logistic regression.

### Introducción

La toma de decisiones es un punto clave en la práctica médica, tanto en el proceso diagnóstico como en el terapéutico. En cualquier situación, estas decisiones deben estar avaladas por criterios de evidencia y experiencia. Para ello,

es preciso recoger y procesar adecuadamente la información referente al estado de salud del paciente (historial, exploración o pruebas diagnósticas, etc.) y confrontarla con la evidencia acumulada en el seguimiento de grupos amplios de pacientes en condiciones controladas. La información disponible se concreta en distintos tipos de variables que reflejan la condición del paciente. A partir de estas variables, es necesario aplicar criterios objetivos que permitan extraer una conclusión adecuada acerca de la posible evolución de la enfermedad, la posibilidad de complicaciones, etc. Considerando la complejidad del problema, la evaluación de la información debe hacerse desde una perspectiva *multivariante*, de manera que se consideren simultáneamente todas las variables implicadas y se obtenga una *generalización* adecuada que permita una clasificación apropiada de nuevos casos.

Las técnicas estadísticas multivariantes proporcionan una solución a este tipo de problemas. Así, el análisis discriminante puede utilizarse en la obtención de un criterio diagnóstico a partir de los valores de varias variables, mientras que el análisis de supervivencia permite evaluar convenientemente la contribución de diversas variables a la supervivencia en distintas circunstancias de interés médico. Por otra parte, el análisis de regresión logística (RL) es adecuado cuando se quiere desarrollar un modelo de predicción de un determinado suceso, en general la probabilidad de complicaciones asociadas a un tratamiento o al estado del paciente.

Todas estas técnicas han conocido un auge muy importante en su utilización en medicina. Este auge se justifica, en parte, por su fácil disponibilidad al estar incluidas en casi todos los paquetes estadísticos de uso habitual. Sin embargo, como sucede con cualquier técnica estadística, su utilización debe tener en cuenta las condiciones apropiadas de aplicación, que en general se referirán a la distribución de las variables con las que se trabaja, la independencia entre ellas, etc.

En el caso particular de la regresión logística, ésta tiene unas condiciones específicas de utilización y unas limitaciones en la interpretación de las conclusiones. Básicamente, la RL parte del supuesto de independencia de las variables y considera un modelo específico de asignación de probabilidades (*modelo logístico*)<sup>1</sup>. Cuando estas condiciones no se cumplen, especialmente debido a la dependencia entre las variables consideradas o a efectos no-lineales no incluidos en el modelo, los resultados de la aplicación de esta técnica son discutibles y pueden estar alejados de la realidad (fig. 1). En estas situaciones, es posible utilizar modelos más elaborados que incluyan interacciones entre las variables y efectos no-lineales. Sin embargo, cuando el problema contiene un elevado número de variables predictoras, su complejidad determina que en la práctica se convierta en un problema difícil de abordar y resolver mediante las técnicas habituales. En este caso, una posible alternativa al empleo de este tipo de análisis basados en técnicas estadísticas, más o me-

Correspondencia: Dr. J. Trujillano.  
Grup de Recerca de Biomatemàtica i Bioestadística.  
Departament de Ciències Mèdiques Bàsiques. Universitat de Lleida.  
Av. Rovira Roure, 44. 25198 Lleida. España.

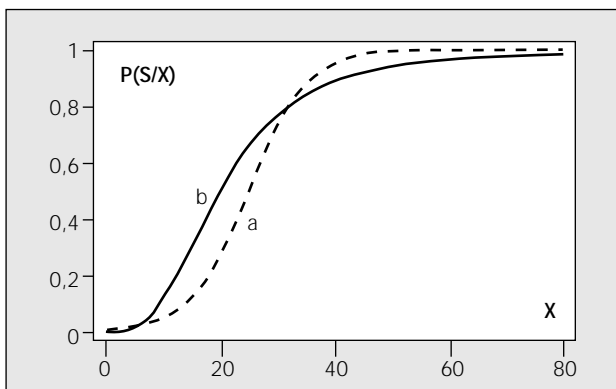


Fig. 1. El modelo logístico. El modelo logístico (a) considera que la función logística representa la relación entre los valores de la variable X y la probabilidad del suceso P(S/X). Si esta relación no se cumple, entonces el método puede proporcionar resultados erróneos. En (b), la P(S/X) se relaciona de manera logística con el Log(X). Si este hecho no se incluye en el modelo, las conclusiones serán erróneas. La RN puede incorporar estas relaciones sin necesidad de formularlas explícitamente.

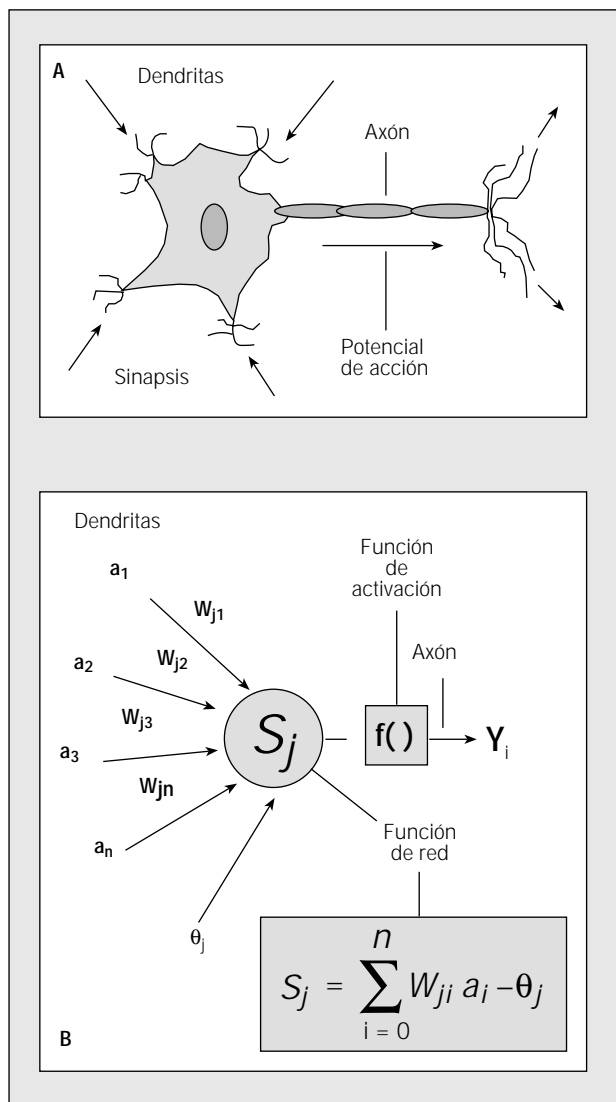
nos clásicas, la encontramos en metodologías propias de otros campos científicos, como puede ser la inteligencia artificial. En particular, las redes neuronales artificiales (RN) son capaces de desarrollar un modelo de predicción que incorpora automáticamente relaciones entre las variables analizadas sin necesidad de incorporarlas explícitamente en el modelo.

El objetivo de esta revisión es proporcionar un acercamiento a la metodología de las RN en el contexto de la medicina, tomando como ejemplo su utilización como alternativa a la RL. Empezaremos presentando las ideas generales acerca de las RN y comentando los procedimientos básicos de desarrollo de una red adecuada para el cálculo de probabilidad de un resultado en función de un conjunto de variables predictoras. Mostraremos un esquema práctico para realizar una aplicación basada en esta técnica, y mediante un ejemplo podremos apreciar algunos de los detalles críticos del proceso de entrenamiento y las ventajas e inconvenientes frente a métodos estadísticos alternativos.

### ¿Qué son las redes neuronales artificiales?

Una RN es un algoritmo de cálculo que se basa en una analogía del sistema nervioso. La idea general consiste en emular la capacidad de aprendizaje del sistema nervioso, de manera que la RN aprenda a identificar un patrón de asociación entre los valores de un conjunto de variables predictoras (*entradas*) y los estados que se consideran dependientes de dichos valores (*salidas*). Desde un punto de vista técnico, la RN consiste en un grupo de unidades de proceso (*nodos*) que se asemejan a las neuronas al estar interconectadas por medio de un entramado de relaciones (*pesos*) análogas al concepto de conexiones sinápticas en el sistema nervioso. A partir de los nodos de entrada, la señal progresa a través de la red hasta proporcionar una respuesta en forma de nivel de activación de los nodos de salida<sup>3</sup>. Los valores de salida proporcionan una predicción del resultado en función de las variables de entrada. Desde el punto de vista de implementación práctica, los nodos son elementos computacionales simples que emulan la respuesta de una neurona a un determinado estímulo. Estos elementos, como las neuronas en el sistema nervioso, funcionan como interruptores: cuando la suma de señales de entrada es suficientemente alta (en el caso de una neurona diríamos que se acumula suficiente neurotransmisor), la neurona manda una señal a las neuronas con las que mantiene contacto (se genera un potencial de acción). Esta situación se modela matemáticamente como una suma de pesos de todas las señales de llegada al nodo que se compara con un umbral característico. Si el umbral se supera, entonces el nodo se dispara, mandando una señal a otros nodos, que a su vez procesarán esa información juntamente con la que reciben de nodos adyacentes (fig. 2). Evidentemente, la respuesta de cada nodo dependerá del valor de las interacciones con los nodos precedentes dentro de la estructura de la red. Como en el caso del sistema nervioso, el poder computacional de una RN deriva no de la complejidad de cada unidad de proceso sino de la densidad y complejidad de sus interconexiones<sup>4</sup>.

La primera implementación práctica de estas ideas se describe en los trabajos de McCulloch y Pitts en 1946. A partir de este punto, algunos de los hitos principales en la investigación de este tipo de técnicas fueron: el diseño por Widrow y Hoff (1961) de la red conocida como *Adalina* (capaz de resolver problemas de regresión lineal), el desarrollo de la red con estructura de *perceptrón simple* en 1959 (con equi-



Figs. 2a y b. Comparación entre la neurona biológica (A) y la neurona artificial (B). (Véase texto.)

$$S_j = \sum_{i=0}^n W_{ji} a_i - \theta_j$$

TABLA 1

**Ejemplos de aplicaciones de redes neuronales artificiales en medicina**

Predicción de resultados Mortalidad/supervivencia Cirugía cardíaca <sup>9</sup> Trasplante de páncreas <sup>10</sup> Cirrosis <sup>11</sup> Pacientes de UCI <sup>12</sup> Neoplasia de colon <sup>13</sup> Neoplasia de páncreas <sup>14</sup> Diagnóstico IAM <sup>15</sup> Nódulo pulmonar <sup>16</sup> Apendicitis <sup>17</sup> Tuberculosis activa <sup>18</sup> Alteración psiquiátrica <sup>19</sup> Otras aplicaciones Análisis de imágenes RMN <sup>20</sup> SPECT <sup>21</sup> Análisis de ondas ECG <sup>22</sup> EEG <sup>23</sup> Farmacología Dosis óptimas <sup>24</sup>
--

TABLA 2

**Clasificación de las redes neuronales artificiales según tipo de aprendizaje y arquitectura**

Aprendizaje	Arquitectura	
	Unidireccional	Realimentada
Supervisado	Perceptrón Adalina Madalina Perceptrón multicapa GRNN LVO	BSB Fuzzy Cog. Map
No supervisado	Máquina de Boltzmann LAM OLAM Mapas de Kohonen Neocognitrón	ART Hopfield BAM
Híbridos	Función de base radial (RBF) Contrapropagación	
Reforzados	Aprendizaje reforzado	

Adalina: *adaptive linear neuron*; Madalina: *many adalinas*; GRNN: *general regression neural network*; LVO: *learning vector quantization*; BSB: *brain state in a box*; LAM: *linear associative memory*; OLAM: *optimal LAM*; ART: *adaptive resonance theory*, y BAM: *bidirectional associative memory*.

valencia al análisis discriminante y regresión logística) y las redes *multicapa* por Rosenblatt en 1986 (que permiten la resolución de situaciones no lineales). Por otra parte, los trabajos teóricos de Bishop y la aportación sobre redes autoorganizadas de Kohonen dotaron de fundamentos formales a este tipo de técnica<sup>5</sup>. A partir de los trabajos pioneros, el interés sobre esta metodología se ha difundido a casi todos los ámbitos de la ciencia. Los distintos aspectos técnicos y las implicaciones de su utilización han sido investigados desde muchos puntos de vista, interesando, entre otros, a matemáticos, físicos, neurólogos, ingenieros, programadores y filósofos. En medicina las áreas de aplicación se han multiplicado rápidamente en la última década (tabla 1)<sup>6-8</sup>, incrementándose continuamente el número de publicaciones que incluyen esta metodología<sup>25</sup>.

Desde un punto de vista práctico, existen muchos tipos de RN. En la tabla 2 se recogen las más características. Para clasificarlas, podemos considerar dos criterios básicos: el modo de aprendizaje y el flujo de información. En una red, el modo de aprendizaje puede ser *supervisado*, es decir, la

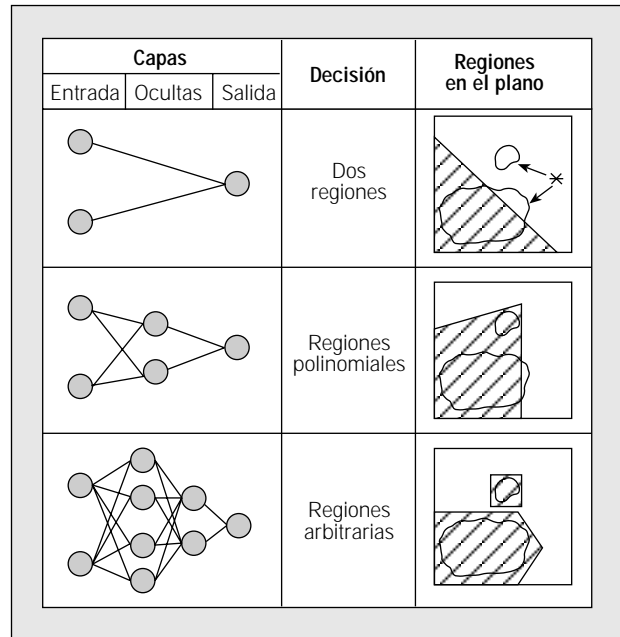


Fig. 3. Capacidad de decisión de las redes neuronales artificiales (perceptrón multicapa). Con 2 variables de entrada, ante un problema de clasificación complejo en el plano (\*) una red sin capa oculta no puede resolverlo; el aumento en el número de capas y nodos ocultos permite encontrar la solución.

red recibe los patrones de entrada y la respuesta observada que debe aprender; o *no supervisado* si la red reconoce automáticamente en los datos el patrón que debe aprender. Por otra parte, el flujo de información que manejan puede ser *unidireccional*, cuando la información sigue una dirección única desde los nodos de entrada a los de salida; o *realimentado*, donde el flujo de información no es único al incorporar circuitos de realimentación entre capas de la red. En una primera aproximación, indicaremos que las redes unidireccionales con aprendizaje supervisado pueden utilizarse para muchos problemas de interés médico.

**El perceptrón multicapa como ejemplo de RN de aplicación en medicina**

Dentro de las redes supervisadas unidireccionales, la estructura más utilizada es el llamado perceptrón multicapa (MLP, *multilayered perceptron*). La arquitectura típica de este tipo de red está constituida por varias capas de nodos con interconexión completa entre ellos. El caso más sencillo en este tipo de red consiste en sólo 2 capas de neuronas, las de entrada y las de salida. De esta manera, podemos obtener un modelo adecuado para problemas lineales del tipo de la regresión lineal múltiple. Si queremos analizar problemas no-lineales, es necesario incorporar otras capas de neuronas intermedias u ocultas (*hidden units*) (fig. 3). En este tipo de red, una neurona recibe distintas entradas y activa una *función de red* (o regla de propagación) con unos pesos de entrada asociados (fig. 2). La computación de estos pesos se sigue de la aplicación de la *función de activación* que determina el nivel de activación de salida de la neurona. La entrada de las neuronas de la primera capa (entrada) son los valores de las variables predictoras y los niveles de activación de las neuronas de la última capa (salida) son los resultados de la red<sup>26</sup>. Dentro de los paráme-

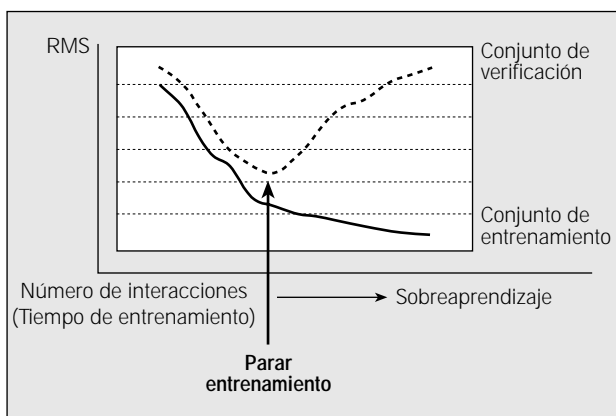


Fig. 4. Criterio de parada del proceso de entrenamiento. RMS: raíz cuadrada del error cuadrático medio.

tros que definen una red, la función de red más utilizada es de tipo lineal, y como función de activación más empleada está la función sigmoidea.

### Proceso de entrenamiento

El entrenamiento consiste en la presentación repetida de un conjunto suficientemente amplio de datos de entrenamiento (*training set*), formado por las entradas y los valores correspondientes de las variables a predecir, hasta conseguir que los pesos internos (interacciones entre nodos) conduzcan a resultados óptimos en la capa de salida, acercándose lo más posible a los resultados esperados. En un contexto médico, el entrenamiento consistiría en presentar a la red, de forma iterativa, los valores de distintas variables clínicas (en forma de valores de la capa de entrada) de cada paciente y conseguir que la red sea capaz de predecir el estado final observado en cada paciente (indicados por el estado de las capas de salida de la red) de la manera más precisa posible.

En la práctica, el ajuste de los pesos durante el entrenamiento se consigue mediante un proceso iterativo cuya finalidad es minimizar una *función de error* que cuantifica la discrepancia entre las predicciones de la red y los valores observados en la muestra. La medida más utilizada para evaluar el error en la predicción (*función de coste*) es la raíz cuadrada del error cuadrático medio (RME, *root-mean-square-error*) entre los valores de salida de la red y sus valores esperados según los datos disponibles. El proceso comienza calculando el RME asociado a la red con los pesos aleatorios de inicio (red no entrenada). Por ejemplo, en una red con sólo 3 capas (una sola oculta), una vez calculado este error, se modifican, de forma retrograda, los pesos de entrada de los nodos de salida (tercera capa) y se calcula para cada neurona oculta (segunda capa) un error próximo. Una vez conocido este error vuelven a actualizarse los pesos de entrada para cada neurona oculta. Este proceso se repite cíclicamente para cada iteración del proceso de entrenamiento. La presentación de los datos en cada ciclo puede hacerse con la totalidad de los mismos, siendo esta estrategia la más habitual, o fraccionando en distintos lotes el conjunto de entrenamiento. Esta última estrategia puede conseguir un mejor aprendizaje en algunas aplicaciones. Observamos que en este tipo de red, el flujo de información es unidireccional (de entrada a salida) pero el flujo de actualización del error es retrógrado (*backpropagation*).

En algunos casos, el proceso de entrenamiento de una red puede llevar a situaciones no deseadas. El riesgo principal consiste en que la red aprenda los ejemplos pero sea incapaz de dar respuestas convenientes en nuevos casos. Para evitar este problema, el proceso de entrenamiento debe considerar otras estrategias adicionales.

### Validación cruzada

Para elaborar una red que sea eficaz es conveniente dividir los datos en 3 conjuntos, atendiendo a que cada uno de ellos mantenga la representatividad de la población origen: a) el conjunto de entrenamiento (*training set*); b) el conjunto de verificación (*test set*), y c) un conjunto de validación (*validation set*).

El conjunto de entrenamiento se usa para ajustar los pesos durante la fase de entrenamiento, mientras que el conjunto de verificación se utiliza para decidir cuándo parar el proceso de entrenamiento. Como criterio general, el entrenamiento debe pararse cuando el error del conjunto de verificación sea mínimo. De esta manera, nos aseguramos que la red es capaz de predecir correctamente los resultados de un conjunto de datos que no forman parte de los ejemplos de entrenamiento. Esta técnica se denomina validación cruzada (*crossvalidation*). Si continuamos el entrenamiento más allá de este punto, la red empieza a aprender de memoria los datos del conjunto de entrenamiento pero pierde capacidad de generalización (fig. 4).

La búsqueda de una generalización óptima, que es la capacidad de la red de proporcionar una respuesta correcta ante patrones que no han sido empleados en su entrenamiento, requiere que se cumplan tres condiciones: a) que la información recogida en las variables sea suficiente —es decir, una selección apropiada de las variables y una buena calidad en la recogida de datos—; b) que la función que aprenda la red sea suave —pequeños cambios en las variables de entrada produzcan pequeños cambios en las variables de salida—, y c) que el tamaño de la base de datos sea suficiente. De esta manera aseguramos que el conjunto de entrenamiento sea representativo de la población a estudio<sup>26</sup>. Excepto la segunda condición, el resto de los requisitos son comunes a cualquier técnica multivariante que se emplee.

Una vez finalizado el entrenamiento, la red (entrenada) evalúa el conjunto de validación y produce las correspondientes predicciones con datos que no se han utilizado en el entrenamiento ni en la validación cruzada. Esta prueba final nos aporta un resultado independiente acerca de la capacidad de generalización de la red.

### Tamaño y arquitectura de la red

La arquitectura de una red viene determinada por el número de capas y nodos que la forman. La complejidad de la red viene determinada por el número de interconexiones que contiene. En general, no es inmediato establecer de forma exacta cuál será la arquitectura ideal para cada aplicación. Así, problemas de discriminación lineal o de regresión logística pueden solucionarse con redes simples. Los problemas surgen al enfrentarse a modelos más complicados (fig. 3). En aplicaciones médicas, un MLP con una única capa oculta puede ser adecuado en muchos casos. Existen *algoritmos evolutivos* que determinan, de forma automática, esta arquitectura óptima al aumentar o retirar nodos o capas del modelo. En cualquier caso, la arquitectura óptima debe alcanzarse, en la práctica, mediante un proceso iterativo, validando la capacidad predictiva de las distintas arquitecturas.

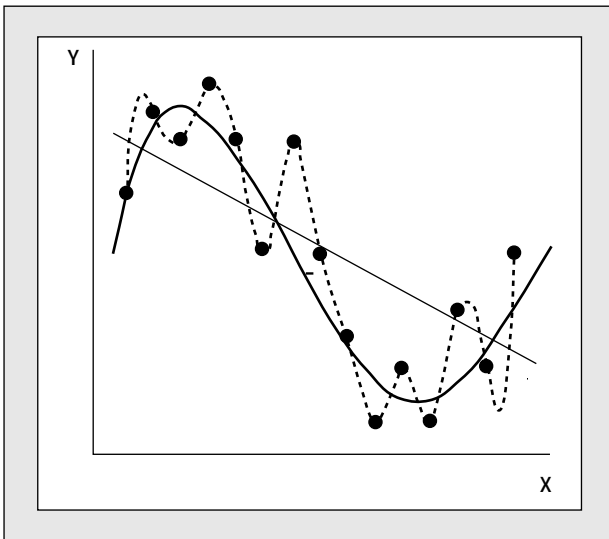


Fig. 5. Búsqueda de la generalización. Relación entre complejidad de arquitectura y resultados. En un problema definido por dos variables (x-y). Los puntos representan los datos. La línea recta es el ajuste con una red sin capa oculta. La línea curva representa el ideal de generalización que se ha conseguido con 3 nodos en la capa oculta. La línea de puntos representa un sobreajuste al utilizar una red con 6 nodos en la capa oculta lo que origina una pérdida de generalización.

tecturas consideradas. Por otra parte, cuanto más compleja sea una red, mayor número de parámetros o pesos deberemos estimar y, por lo tanto, necesitará mayor número de patrones para ser entrenada de manera adecuada.

En el otro extremo, la utilización de una red demasiado compleja para solucionar un problema sencillo nos conduce a un *sobreajuste* (*overfitting*) que dificulta la capacidad de generalización de la red (fig. 5). Como regla general, para reducir el número de parámetros de una red es conveniente seleccionar apropiadamente las variables de entrada, descartando variables poco informativas. Sin embargo, esta selección no es tan sencilla como en los métodos multivariantes habituales y puede requerir distintas etapas de entrenamiento. Desde un punto de vista más técnico, existen procedimientos propios de la metodología de redes que simplifican la estructura de la red. A modo de ejemplo, algunos de estos métodos consisten en compartir pesos entre varios nodos (*weight sharing*), realizar un podado de la red (*pruning*) eliminando los pesos con menor influencia en el resultado del modelo final o aplicar el método de decaimiento de pesos (*weight decay*), eliminando automáticamente los pesos que tienden a cero<sup>26</sup>.

**Correspondencia entre redes neuronales artificiales y técnicas estadísticas**

En el ámbito de la medicina, la utilización de las RN se ha desarrollado paralelamente a su comparación con técnicas estadísticas. Dependiendo del problema específico estudiado, esta confrontación ha llevado, durante esta última década, a alternar entre el optimismo<sup>27,28</sup> y el pesimismo<sup>29,30</sup> en la utilización de las redes en el entorno de los estudios médicos.

La tabla 3 recoge la correspondencia entre algunas redes y distintos procedimientos estadísticos habitualmente utilizados en medicina. Es interesante apreciar que existen algunos tipos de redes que no poseen una correspondencia

TABLA 3

**Relación entre redes neuronales artificiales y técnicas estadísticas**

Red neuronal artificial	Técnica estadística
Perceptrón simple (nodo umbral)	Análisis discriminante
Perceptrón simple (nodo sigmoideo)	Regresión logística
Adalina	Regresión lineal
Perceptrón multicapa	Regresión no lineal
Aprendizaje no supervisado	Análisis de componentes principales
Mapas de Kohonen	Escalas multidimensionales
LVQ (cuantificación de vectores)	Análisis discriminante
Función de base radial (RBF)	Método de regresión de Kernel
Sin relación	
Contrapropagación	
Redes ART	
Modelo de Hopfield	

TABLA 4

**Ventajas y desventajas de las redes neuronales artificiales (perceptrón multicapa) frente a técnicas de regresión logística**

<p>Ventajas de las redes</p> <ul style="list-style-type: none"> <li>Necesitan menos formalismo estadístico para su desarrollo</li> <li>Detectan relaciones no lineales</li> <li>Detectan interacciones entre variables predictoras</li> <li>Tienen múltiples algoritmos de entrenamiento</li> </ul> <p>Desventajas de las redes</p> <ul style="list-style-type: none"> <li>Son «cajas negras» para identificar las interacciones</li> <li>Tienen dificultad de utilización</li> <li>Necesitan mayores recursos computacionales</li> <li>Metodología menos conocida</li> <li>Implementación llena de procesos empíricos</li> <li>Problemas metodológicos no resueltos totalmente</li> </ul>
--

concreta con un método estadístico<sup>31</sup>. La comparación más frecuentemente analizada en la literatura se realiza entre el tipo de red más empleada (MLP + *backpropagation*) y la regresión logística múltiple<sup>32,33</sup>. En la tabla 4 se muestran algunos puntos destacando las ventajas y desventajas de las redes de acuerdo con esta comparación. En una revisión de 28 aplicaciones distintas a partir de la bibliografía, Sargent concluye que las redes son, en el peor de los casos, equivalentes o en general ligeramente superiores a la regresión logística múltiple al no tener que depender de exigencias rígidas de independencia de las variables o de los supuestos inherentes al modelo logístico<sup>34</sup>.

En una aplicación concreta, las redes pueden interpretar de manera distinta la información contenida en las variables respecto a cómo se interpreta esta información en un procedimiento estadístico. Esto nos obliga a analizar cuidadosamente la distinta contribución de cada variable al modelo final y a interpretar sus interdependencias<sup>35-37</sup>. A partir de este análisis es posible mejorar los modelos estadísticos (por ejemplo, añadiendo interacciones encontradas entre las variables). De este modo, ambas técnicas pueden colaborar para proporcionar un modelo final adecuado al problema objeto de estudio.

**Indicaciones prácticas acerca de la aplicación de una red neuronal artificial en el entorno de predicción de resultados en medicina (cálculo de una probabilidad)**

El desarrollo de una RN necesita planificarse adecuadamente para conseguir una red convenientemente entrenada que alcance una precisión óptima<sup>38</sup>. A modo indicativo, en la figura 6 se muestra un esquema básico de actuación y en

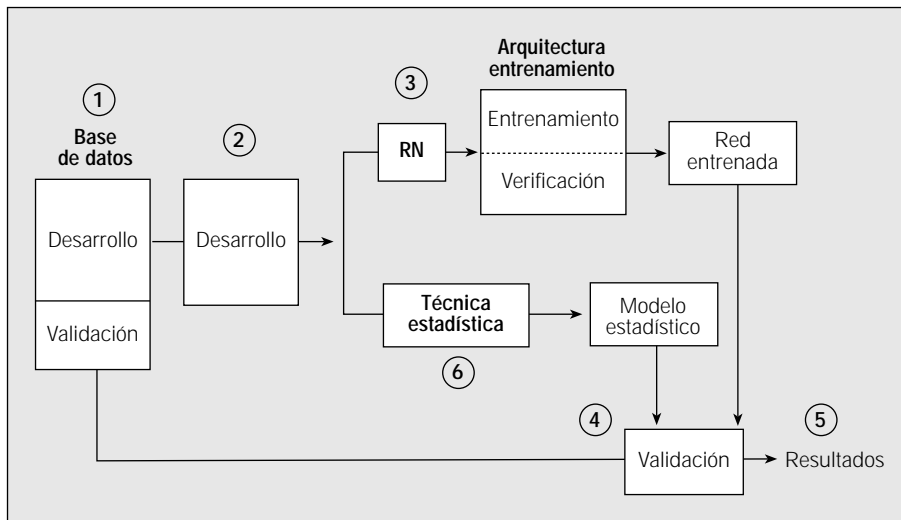


Fig. 6. Secuencia del desarrollo de una aplicación basada en red neuronal artificial. RN: red neuronal artificial. (Véase texto.)

TABLA 5

**Principales errores cometidos en las publicaciones con aplicaciones basadas en redes neuronales artificiales (predicción de resultados)**

Poca o nula descripción del tipo, arquitectura y proceso de entrenamiento de la red
Tamaño del conjunto de verificación ( <i>test set</i> ) inapropiado o poco representativo
Sobreajuste ( <i>overfitting</i> ) por estructura muy compleja de la red (en nodos y capas). La red ajusta funciones implausibles
No utilizar el conjunto de validación ( <i>validation set</i> ). No se expone la red a datos no utilizados para su entrenamiento
No comparar la red neuronal con la técnica estadística adecuada

la tabla 5 se señalan, de acuerdo con el trabajo de Schwarzer et al, los errores más frecuentes cometidos en trabajos publicados que no han seguido esta planificación<sup>39</sup>. Para facilitar una aproximación práctica al uso de esta metodología, planteamos un ejemplo sencillo basado en datos reales con el que seguiremos los distintos pasos. Supongamos que queremos identificar a los pacientes que desarrollarán fracaso renal (definido con criterios estándar) durante su ingreso en una unidad de cuidados intensivos. Para ello, decidimos analizar un modelo que utiliza 3 variables predictoras (simplificamos el problema para hacerlo más didáctico). Las variables consideradas son la *edad*, el peor valor de *tensión arterial media* en su primer día de estancia y la identificación de una *infección* dentro del diagnóstico de entrada. Sabemos que en nuestra serie es más frecuente tener fracaso renal si el paciente tiene más edad, menos presión arterial y diagnóstico de infección. Pero también sospechamos que las variables no son independientes, ya que los pacientes con infección tienden a tener una presión arterial menor, algunos grupos diagnósticos (definidos por su edad) presentan un mayor número de infecciones y la edad condiciona algunos aspectos de la presión arterial. Por lo tanto, debemos buscar un modelo que puede recoger estas relaciones. La variable de resultado será la probabilidad de tener fracaso renal.

*Paso 1: base de datos adecuada*

El resultado obtenido con una RN depende de los datos que se utilizan para su entrenamiento y, por lo tanto, los sesgos

de muestreo pueden influir negativamente el resultado. Aunque se han conseguido buenos resultados con series muy grandes (80.606 pacientes)<sup>40</sup> o muy pequeñas (sólo 74 casos)<sup>41</sup>, se recomienda utilizar, como mínimo, 5-10 observaciones por parámetro estimado, es decir por cada una de las conexiones de la red. Por ejemplo, una red con 17 variables de entrada, 9 nodos en la capa oculta y un nodo de salida (todos plenamente interconectados) tiene 162 parámetros (son 17 por 9 más 9) que exigen un conjunto de entrenamiento mayor de 800 casos<sup>42</sup>. Para nuestro ejemplo disponemos de una base de datos con 1.000 pacientes, de los que 310 presentaron fracaso renal.

*Paso 2: conjuntos de entrenamiento, verificación y validación*

La partición de la serie de datos en los conjuntos de desarrollo (entrenamiento y verificación) y el correspondiente conjunto de validación, determina que el tamaño muestral sea suficientemente grande. En aquellos casos en que no sea así, se ha propuesto la utilización de técnicas de remuestreo (*bootstrap*) con lo que se consigue tener múltiples conjuntos de entrenamiento que aseguran, dentro de las posibilidades y limitaciones de estas técnicas, un proceso de entrenamiento adecuado que conducirá a una buena generalización<sup>43</sup>. En nuestro ejemplo sobre fracaso renal, la división (de forma aleatoria) aporta 400 pacientes para entrenamiento, 300 para verificación y 300 para validación.

*Paso 3: construcción y entrenamiento de la red*

Las características de las redes neuronales determinan que su utilización requiera de programas informáticos adecuados. En este punto, existen múltiples opciones comerciales y de libre distribución<sup>44</sup>. Para una referencia actualizada al respecto, el lector interesado puede consultar la página <ftp://ftp.sas.com/pub/neural/FAQ.html>. En nuestro caso, utilizamos el programa Qnet (Vesta Services Inc). Usando estos programas, podemos empezar a probar qué arquitectura es más conveniente. Ya hemos comentado que este proceso es básicamente empírico. Podemos citar ejemplos que funcionan con estructuras muy simples (19 nodos de entrada/2 nodos ocultos/1 nodo de salida)<sup>45</sup> o muy complejas

(16/35/10/1)<sup>12</sup>. En cualquier caso, en la publicación de resultados deben describirse todos los parámetros de construcción (capas, nodos, tipo de interconexiones) y del proceso de entrenamiento (tipo, número de iteraciones, coeficiente de aprendizaje, momento, etc.).

Siguiendo con nuestro ejemplo, utilizamos un perceptrón multicapa entrenado con algoritmo de *backpropagation* con 3 capas, seleccionando como arquitectura óptima una estructura (3/3/1) con plena interconexión. La función de activación es sigmoidea, parando el entrenamiento, según el criterio de validación cruzada a las 1.500 iteraciones. De acuerdo con esta arquitectura, debemos ajustar 12 parámetros, con lo que los 400 casos del conjunto de entrenamiento son suficientes de acuerdo con el criterio general.

**Paso 4: validación de la red**

Debe comprobarse la capacidad de generalización de la red enfrentándola a datos distintos de los utilizados en su entrenamiento (conjunto de validación). En este ejemplo, disponemos de 300 casos en el conjunto de validación.

**Paso 5: evaluación de los resultados (precisión de la red)**

En este ejemplo se trata de predecir la probabilidad de fracaso renal. En este tipo de problemas, donde la predicción es una probabilidad, lo indicado es evaluar la *discriminación* y la *calibración* de la red<sup>46,47</sup>. Para valorar la discriminación (capacidad de distinguir entre dos estados) se emplean tablas de contingencia eligiendo punto de corte (normalmente 0,5), estableciendo porcentajes de correcta clasificación y analizando las curvas ROC resultantes (especialmente calculando el área bajo la curva ROC)<sup>48,49</sup>. Para comprobar la calibración (exactitud del modelo comparando la probabilidad esperada y la observada) se utilizan las pruebas de bondad de ajuste de Hosmer-Lemeshow<sup>50</sup> que permiten evaluar las curvas de calibración correspondientes a los resultados del método.

**Paso 6: comparación de los resultados entre redes neuronales artificiales y regresión logística**

En nuestro ejemplo, comparamos el resultado de la red con los correspondientes a un modelo de regresión logísti-

ca múltiple que utiliza las 3 variables indicadas anteriormente. Al comparar los resultados (en este caso sólo medimos la propiedad de discriminación con tablas de contingencia y curvas ROC) (fig. 7), podemos comprobar que la red consigue resultados más apropiados. Estos mejores resultados de la red tienen como contrapartida una necesidad de utilizar más parámetros (12) que la regresión logística (3) (fig. 8).

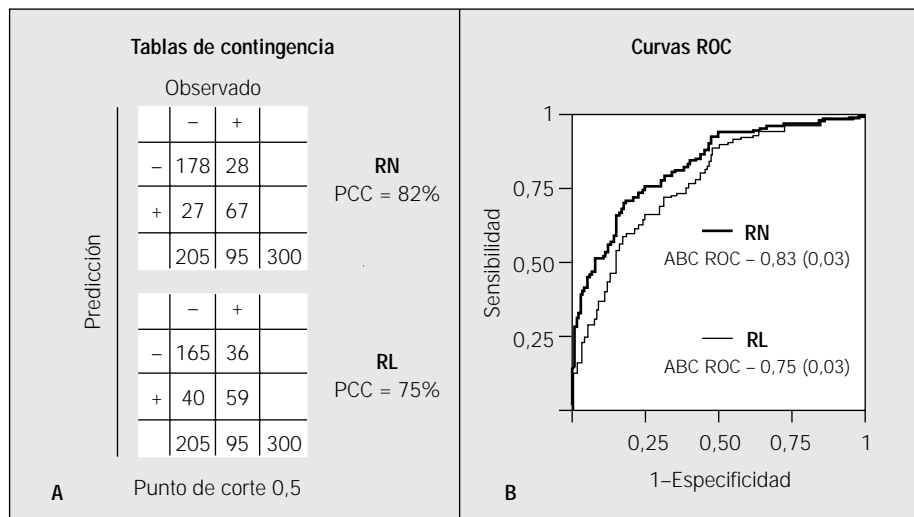
Para que la RL se aproxime a los resultados de la RN en este ejemplo, debemos añadir algunos términos de interacción para tener en cuenta la dependencia entre las variables. Sin embargo, es difícil decidir qué términos deben ser incluidos. Si el problema implica más variables, este escollo puede ser difícilmente superable.

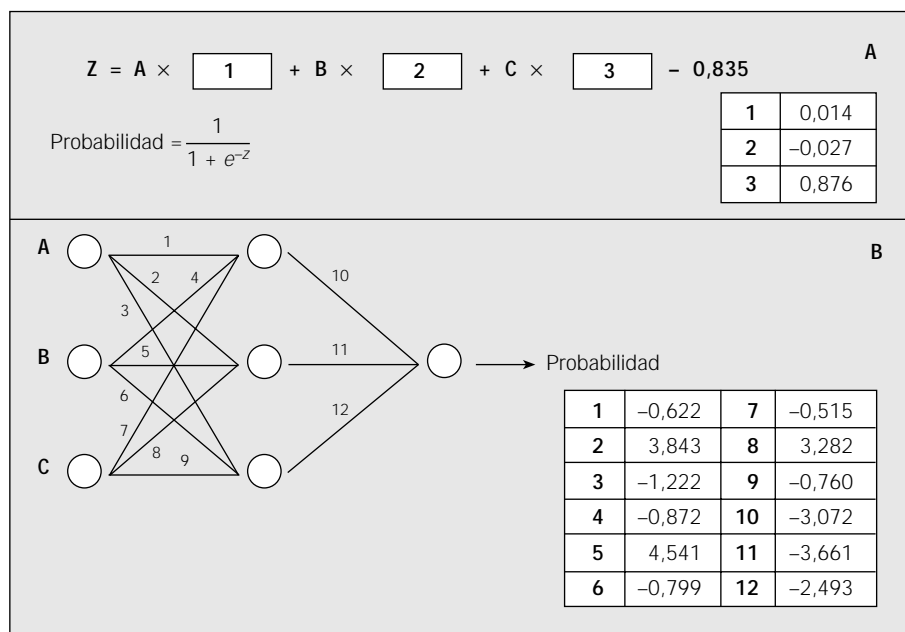
**Conclusiones**

Las RN proporcionan un método general para desarrollar modelos de predicción en medicina. La ventaja principal de esta técnica, si se aplica convenientemente, radica en su capacidad para incorporar interacciones entre las variables sin necesidad de incluirlas *a priori*. Además, su aplicación no queda restringida a un tipo determinado de distribución de los datos. Como principal desventaja, sin embargo, debemos indicar que proporciona un modelo que es esencialmente una caja negra. La RN es capaz de predecir resultados, pero no disponemos de una interpretación evidente de los parámetros en los mismos términos en que podemos interpretar los resultados de una RL.

En esta revisión hemos considerado fundamentalmente el modelo de red más sencillo. En la actualidad se trabaja con nuevos tipos de redes más potentes y que incorporan técnicas adicionales como son los algoritmos genéticos y los modelos híbridos<sup>52-54</sup>. Con ello, se empieza a disponer de herramientas que se fundamentan en el cálculo intensivo y que desafían a los planteamientos estadísticos convencionales. A nuestro entender, lejos de representar una amenaza, estas técnicas proporcionan nuevos puntos de vista que pueden ayudar a obtener herramientas más eficaces en muchas aplicaciones prácticas. Así, el futuro no debería plantearse en términos de competencia entre estos nuevos métodos y la estadística, sino que debería contemplar su acercamiento y complementación para construir modelos de predicción que sean más válidos.

*Figs. 7a y b. Resultados de la comparación entre un modelo basado en red neuronal artificial (RN) y otro basado en regresión logística múltiple (RL). Aplicado al conjunto de validación del ejemplo seguido en el texto con 300 pacientes. A: tablas de contingencia donde se evalúa como resultado positivo (+) presentar un fracaso renal; PCC: porcentaje de correcta clasificación. B: curvas ROC; ABC ROC: área bajo la curva ROC (error típico).*





Figs. 8a y b. Número de parámetros. En nuestro ejemplo, el modelo de regresión logística (A) sólo precisa de 3 parámetros (coeficientes de la ecuación logística) para el cálculo de la probabilidad de presentar fracaso renal. La red neuronal (B) ha conseguido mejores resultados, pero necesita 12 parámetros (pesos de las conexiones). (A): edad; (B): valor de presión arterial media, y (C): presencia de infección.

**REFERENCIAS BIBLIOGRÁFICAS**

1. Steyerberg EW, Eijkemans MJC, Harrell FE, Habbema JDF. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Statist Med* 2000;19:1059-79.
2. Burke HB. Artificial neural networks for cancer research: outcome prediction. *Semin Surg Oncol* 1994;10:73-9.
3. Cross BS, Harrison RF, Kennedy RL. Introduction to neural networks. *Lancet* 1995;346:1075-9.
4. Sarle WS. Neural network FAQ, part 1 of 7: introduction, periodic posting to the Usenet newsgroup comp.ai.neural-nets, URL: ftp://ftp.sas.com/pub/neural/FAQ.html, 1997.
5. Bishop CM. Neural networks for pattern recognition. Oxford: Oxford University, 1995.
6. Baxt WG. Application of artificial neural networks to clinical medicine. *Lancet* 1995;346:1135-8.
7. Armoni A. Use of neural networks in medical diagnosis. *MD Computing* 1998;15:100-4.
8. Dybowski R, Gant V. Artificial neural networks in pathology and medical laboratories. *Lancet* 1995;346:1203-7.
9. Orr RK. Use of a probabilistic neural network to estimate the risk of mortality after cardiac surgery. *Med Decis Making* 1997;17:178-85.
10. Dorsey SG, Waltz CF, Brosch L, Connerney I, Schweitzer EJ, Barlett ST. A neural network model for predicting pancreas transplant graft outcome. *Diabetes Care* 1997;20:1128-33.
11. Lapuerta P, Rajan S, Bonacini M. Neural networks as predictors of outcomes in alcoholic patients with severe liver disease. *Hepatology* 1997;25:302-6.
12. Wong LSS, Young JD. A comparison of ICU mortality prediction using the APACHE II scoring system and artificial neural network. *Anaesthesia* 1999;54:1048-54.
13. Snow PB, Kerr DJ, Brandt JM, Rodvold DM. Neural network and regression predictions of 5-year survival after colon carcinoma treatment. *Cancer* 2001;91:1673-8.
14. Montie JE, Wei JT. Artificial neural networks for prostate carcinoma risk assessment. An overview. *Cancer* 2001;91:1647-52.
15. Ellenius J, Groth T, Lindahl B, Wallentin L. Early assessment of patients with suspected acute myocardial infarction by biochemical monitoring and neural network analysis. *Clin Chem* 1997;43:1919-25.
16. Nakamura K, Yoshida H, Engelmann R, MacMahon H, Katsuragawa S, Ishida T, et al. Computerized analysis of the likelihood of malignancy in solitary pulmonary nodules with use of artificial neural networks. *Radiology* 2000;214:823-30.
17. Eberhart RC, Dobbins RW, Hutton LV. Neural network paradigm comparisons for appendicitis diagnosis. En: Proceedings of the fourth annual IEEE Symposium on Computer-based Medical Systems. Los Alamitos, CA: IEEE Computer Society Press, 1991; p. 298-304.
18. El-Solh AA, Hisao C, Goodnough S, Serghani J, Grant BJB. Predicting active pulmonary tuberculosis using an artificial neural network. *Chest* 1999;116:968-73.
19. Somoza E, Somoza JR. A neural-network approach to predicting admission decisions in a psychiatric emergency room. *Med Decis Making* 1993;13:273-80.
20. Hare BJ, Prestegard JH. Application of neural networks to automated assignment of NMR spectra of proteins. *J Biomol NMR* 1994;4:35-46.
21. Kippenhan JS, Barker WW, Angel J, Grady C, Duara R. Neural-networks classification of normal and Alzheimer's disease subjects using high-resolution and low-resolution PET cameras. *J Nucl Med* 1994;35:7-15.
22. Edenbrandt L, Devine B, Macfarlane PW. Classification of electrocardiographic ST-T segments-human expert vs artificial neural network. *Eur Heart J* 1993;14:464-8.
23. Anderer P, Saletu B, Kloppel B, Semlitsch HV, Werner H. Discrimination between demented patients and normal based on topographic EEG slow wave activity: comparison between z statistics, discriminant analysis and artificial neural network classifiers. *Electroencephalogr Clin Neurophysiol* 1994;91:108-17.
24. Yamamura S, Nishizawa K, Hirano M, Momose Y, Kimura A. Prediction of plasma levels of aminoglycoside antibiotic in patients with severe illness by means of an artificial neural network simulator. *J Pharm Sci* 1998;1:95-101.
25. Dayhoff JE, DeLeo JM. Artificial neural networks. Opening the black box. *Cancer* 2001;91:1615-35.
26. Martín del Brio B, Sanz Molina A. Redes neuronales y sistemas borrosos. Madrid: RA-MA, 1997.
27. Eldar S, Siegelmann HT, Buzaglo D, Matter I, Cohen A, Sabo E, et al. Conversion of laparoscopic cholecystectomy to open cholecystectomy in acute cholecystitis: artificial neural networks improve the prediction of conversion. *World J Surg* 2002;26:79-85.
28. Duh MS, Walker AM, Pagano M, Kronlund K. Prediction and cross-validation of neural networks versus logistic regression: using hepatic disorders as an example. *Am J Epidemiol* 1998;147:407-13.
29. Borque A, Sanz G, Allepuz C, Plaza L, Rioja LA. The use of neural networks and logistic regression analysis for predicting pathological stage in men undergoing radical prostatectomy: a population based study. *J Urol* 2001;166:1672-8.
30. Sinha M, Kennedy CS, Ramundo ML. Artificial neural network predicts CT scan abnormalities in paediatric patients with closed head injury. *J Trauma* 2001;50:308-12.
31. Sarle WS. Neural networks and statistical models. Proceedings of the 19th annual SAS Users Group International Conference, Cary, NC (SAS Institute), 1994:1538-50.
32. Tu JV, Guerriere MRJ. Use of a neural network as a predictive instrument for length of stay in the intensive care unit following cardiac surgery. *Comput Biomed Res* 1993;26:220-9.
33. Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol* 1996;49:11:1225-31.
34. Sargent DJ. Comparison of artificial neural networks with other statistical approaches. Results from medical data sets. *Cancer* 2001;91:1636-42.
35. Lang EW, Pitts LH, Damron SL, Rutledge R. Outcome after severe head injury: an analysis of prediction based upon comparison of neural network versus logistic regression analysis. *Neuro Res* 1997;19:274-80.
36. DiRusso SM, Sullivan T, Holly C, Nealon S, Savino J. An artificial neural network as a model for prediction of survival in trauma patients: validation for a regional trauma area. *J Trauma* 2000;49:212-23.



37. Selker HP, Griffith JL, Patil S, Long WJ, D'Agostino RB. A comparison of performance of mathematical predictive methods for medical diagnosis: identifying acute cardiac ischemia among emergency department patients. *J Investig Med* 1995;43:468-76.
38. Schwarzer G, Vach W, Schumacher M. On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology. *Statist Med* 2000;19:541-61.
39. Levine RF. Conference concluding remarks. *Cancer* 2001;91:1696-7.
40. Lippmann RP, Shahian DM. Coronary artery bypass risk prediction using neural networks. *Ann Thorac Surg* 1997;63:1635-43.
41. Guh JY, Yang CY, Yang JM, Chen LM, Lai YH. Prediction of equilibrated postdialysis BUN by an artificial neural network in high-efficiency hemodialysis. *Am J Kidney Dis* 1998;31:638-46.
42. Clermont G, Angus DC, DiRusso SM, Griffin M, Linde-Zwirble WT. Predicting hospital mortality for patients in the intensive care unit: a comparison of artificial neural networks with logistic regression models. *Crit Care Med* 2001;29:291-6.
43. Tourassi GD, Floyd CE. The effect of data sampling on the performance evaluation of artificial neural networks in medical diagnosis. *Med Decis Making* 1997;17:186-92.
44. Sarle WS. Neural Network FAQ, part 5 and 6 of 7: free software and commercial software. Disponible en: URL: <ftp://ftp.sas.com/pub/neural/FAQ.html>, 1997.
45. Doyle HR, Dvorchik I, Mitchell S, Marino JR, Ebert FH, McMichael J, et al. Predicting outcomes after liver transplantation. A connectionist approach. *Ann Surg* 1994;219:408-15.
46. Concato J. Challenges in prognostic analysis. *Cancer* 2001;91: 1607-14.
47. Van Houwelingen HC. Validation, calibration, revision and combination of prognostic survival models. *Statist Med* 2000;19:3401-15.
48. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29-36.
49. Burgueño MJ, García-Bastos JL, González-Buitrago JM. Las curvas ROC en la evaluación de las pruebas diagnósticas. *Med Clin (Barc)* 1995; 104:661-70.
50. Lemeshow S, Hosmer DW. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 1982; 115:92-106.
51. Jefferson MF, Pendleton N, Lucas SB, Horan MA. Comparison of a genetic algorithm neural network with logistic regression for predicting outcome after surgery for patients with non small cell lung carcinoma. *Cancer* 1997;79:1338-42.
52. Dybowski R, Weller P, Chang R, Gant V. Prediction of outcome in critically ill patients using artificial neural network synthesised by genetic algorithm. *Lancet* 1996;347:1146-50.